Multiple Comparison among Groups of Growth Curves

by Toshinari Kamakura[1] and Tsuyoshi Takizawa[2]

The problem of comparing a sequence of independent experiments divided into several groups with a control is discussed under the logistic growth-curve models. We propose a method for constructing multiple testing procedures using the closed testing procedures and the random-effect model for summarizing estimated values of parameters.

Introduction

In toxicological studies it is of great importance to evaluate side effects or toxicity of new drugs, industrial compounds, and environmental contaminants. We are interested in investigating the maximum dose levels below which no toxicity is observed or, if a maximum level exists, if it is tolerable. We call this level the maximum noneffective dose (MNED). Ruberg (*1*) considers the problems of inference about the minimum effective dose (MED) by comparing various dose groups with a control and reaches the conclusion that the simulation studies have superior contrast procedures compared to the other multiple comparison procedures (*2,3*). If we design an experiment in which differences in doses among groups are small, MNED and MED are almost the same. In this article we focus on inferring the MNED from the viewpoint of a safe dose.

In the usual dose–response studies we can use the standard multiple comparison techniques as described above. However, it is difficult to apply them in cases where the responses are continually observed with constant exposure to some dose level and when the observations are arranged on the time axis for each experiment. The problem is how we can select the time point at which to compare the two groups. Even if we select one point for comparison, we may lose other information such that the observations of one group are always fewer than those of the other groups. This situation is illustrated in the Figure 1. We are interested in comparing groups in which each animal is observed continually, and we propose a multiple comparison technique for these types of data sets.

Growth-Curve Modeling and Maximum Noneffective Dose

Suppose we have the following data sets:

$$\{x_{ijk} \; ; i = 0, 1, .., a; j = 1, ..., n; k = 1, ..., T\}$$

[1]Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112, Japan.
[2]Nippon Roche K. K., 200 Kajiwara, Kamakura 247, Japan.
Address reprint requests to T. Kamakura, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112, Japan.
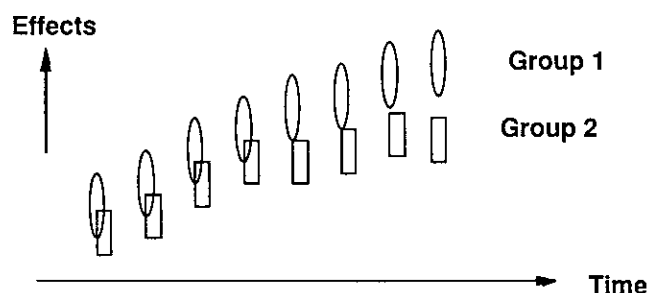
FIGURE 1. Difficulty in comparing data sets arranged on a time axis.

The subscripts i, j, and k mean, respectively, the ith group, jth individual in each group, and kth observation for the each individual on the time axis. For example, x_{ijk} is the weight of the jth animal at time k in the ith group. We assume zero dose for $i = 0$ as a control group and that dose level increases monotonously with index i.

Our main purpose is to compare the weight curves observed on the time axis for each animal in the $(a + 1)$ groups. We would like to know whether there are any differences in weight curves among dose groups. Can we draw the line that discriminates the dose level below which we may conclude no effects or no obstructions in growths and compare it to the control group? This question drives us to investigate the maximum noneffective dose statistically. Yoshimura (*4*) considers this problem with no time-dependent observations from the viewpoint of the standard multiple comparisons. However, we cannot extend those multiple comparison techniques to the problem of weight curves. First we will fit some growth curve model with a few parameters and second we will compare the derived estimates of these parameters.

We assume the logistic growth curves as weight curves for animals, which is described as follows:

$$x_j = \frac{K}{1 + e^{-\beta(t_j + \mu)}} + \epsilon_i \quad (j = 1, ..., T) \qquad (2)$$

Here ϵ is independently normally distributed with mean 0 and variance σ^2. Note that the indexes i and k are omitted for

simplicity. The parameter K is interpreted as the final weight of an animal with constant exposure to a dose of some level. So the estimate of K is one of our primary interests. We first compute the estimates by maximizing the likelihood,

$$L = \prod_{j=1}^{T} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2} \left\{ x_j - \frac{K}{1 + e^{-\beta(t_j+\mu)}} \right\}^2 \right] \quad (3)$$

for each individual. In fact, we must solve the $n(a+1)$ likelihood equations. Then we would obtain $4n(a+1)$ parameter estimates,

$$\left\{ \hat{K}_{ij}, \hat{\beta}_{ij}, \hat{\mu}_{ij}, \hat{\sigma}_{ij}^2 \right\} (i = 0, 1, ..., a; \ j = 1, ...n)$$

and their asymptotic variances. In the following section we propose a method to summarize these estimates for multiple comparisons.

Multiple Comparisons by Random-Effects Model

It is useful to use a random-effect model for summarizing estimated variables for each group. Korn and Whittemore (5) use the random-effect model for obtaining overall estimates assuming a multiple logistic model for each patient.

We assume that $\hat{K}_{ij}$ is normally distributed with mean K_{ij} and variance σ_{ij}^2 and that K_{ij} and η_i^2 are obtained by solving the following likelihood equations:

$$\begin{cases} \sum_{j=1}^{n} \frac{\hat{K}_{ij} - K_i}{\sigma_{ij}^2 + \eta_i^2} = 0 \\ \sum_{j=1}^{n} \frac{(\hat{K}_{ij} - K_i)^2}{(\sigma_{ij}^2 + \eta_i^2)^2} - \sum_{j=1}^{n} \frac{1}{\sigma_{ij}^2 + \eta_i^2} = 0 \end{cases} \quad (4)$$

In these equations we can use the asymptotic variances of $\hat{K}_{ij}$, $\hat{\sigma}_{ij}^2$, as the variances σ_{ij}^2. The asymptotic variances and covariances of $\hat{K}_i$ and $\hat{\eta}_i^2$ are calculated as follows:

$$\begin{cases} Avar(\hat{K}_i) = \left(\sum_{j=1}^{n} \frac{1}{\sigma_{ij}^2 + \hat{\eta}_i^2} \right)^{-1} \\ Avar(\hat{\eta}_i^2) = \left(\frac{1}{2} \sum_{j=1}^{n} \frac{1}{(\sigma_{ij}^2 + \hat{\eta}_i^2)^2} \right)^{-1} \\ ACov(\hat{K}_i, \hat{\eta}_i^2) = 0 \end{cases} \quad (5)$$

Now we would like to obtain the MNED level by multiply comparing the estimates $\{\hat{K}_i \ i = 0, \ ..., a\}$. First we consider the problem of the hypothesis testing where the null hypothesis is

$$H_0 \ : \ K_0 = K_1 = \cdots = K_a \ .$$

We define that

$$r' = (K_0, K_1, \cdots, K_a).$$

Then the null hypothesis is expressed as follows:

$$H_0 \ : \ Cr = o,$$

where the matrix $C [a \times (a+1)]$ is as follows:

$$C = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}$$

We consider the following test statistic.

$$T = (C\hat{r})' \left\{ C\mathcal{D}(\hat{r})C' \right\}^{-1} (C\hat{r}) \quad (6)$$

The operator $\mathcal{D}$ is used to generate the covariance matrix of the vector argument. In our case the covariance matrix is a simple, diagonal matrix whose elements correspond to the asymptotic variances. Under the null hypothesis, the statistic given by Equation 6 has asymptotically a chi-square distribution with a degrees of freedom. For multiple comparison we can use the closed-testing procedures proposed by Marcus et al. (6), which require that sets of hypotheses are closed under intersection and that each test is of level α. Then we can assure that the overall error rate is less than α if we use these multiple comparison procedures.

We consider the following set of hierarchical hypotheses closed under intersection:

$$\begin{aligned} H_1 &: K_0 = K_1 \\ H_2 &: K_0 = K_1 = K_2 \\ &\vdots \\ H_a &: K_0 = K_1 = K_2 = \cdots = K_a \end{aligned}$$

The closed procedures are constructed by testing each null hypothesis with a level α and finding the hypothesis $H_{\lambda 0}$; H_λ for $\lambda < \lambda_0$ is not rejected and H_λ for $\lambda > \lambda_0$ is rejected with a level α. Here we can use the test statistic (Eq. 5) for each test.

Examples

We consider two examples in which a new drug is tested for toxicity. The first data are male body weights (grams) for a control and a dose group (200 mg/kg) from a 5-week toxicity study in rats. The summary statistics are shown in Table 1.

What troubles toxicologists is that the control versus dose group comparison shows significant difference at days 4–32 and but is not significant at day 35 by the t-test. Should they conclude significance or nonsignificance? We estimate the final body

Table 1. Summary of statistics of control and dose groups.

Group		Day										
		0	4	7	11	14	18	21	25	28	32	35
Control	Mean	121.1	158.4	186.0	218.6	241.1	273.6	293.4	320.8	334.8	355.7	361.4
($n = 9$)	SE	1.3	2.4	2.5	3.6	4.2	5.2	6.2	7.0	7.5	8.8	9.7
Dose	Mean	122.4	152.4	173.4	203.0	225.5	255.0	275.9	297.4	309.0	330.4	337.8
($n = 10$)	SE	1.5	1.6	2.0	2.5	3.3	3.6	4.2	5.3	5.2	6.4	6.6

weights K^{values} in the random-effects model. The mean K is 402.20 g for the control and 382.72 g for the dose group. The test statistic of Equation 6 with $a = 1$ results in 1.53, and therefore we can conclude that the two groups are not significantly different from the viewpoint of the final estimated body weights compared to the upper probability of chi-squared distribution with 1 degree of freedom.

Second, we choose for illustration the data set consisting of a control group and four dose groups: for each group, weights of 16 rats were taken at 27 time points. Dose groups levels are 15, 35, 85, and 200 mg/kg. Figure 2 shows the plots of growth of 16 rats in 15 mg/kg dose group.

The estimate of the final body weight and its asymptotic variance in each group are given in Table 2. Table 3 shows the values of statistics obtained from Equation 6. From the table, we can find that the dose group of 200 mg/kg is significantly different from the control group in the final body weight.

Sometimes laboratory workers make an error in administering drugs to rats and hurt the rats' throats. Subsequently the rats

Table 2. Estimated final weights (grams) and their asymptotic variances.

Group number	$\hat{K}_i$	$A\mathit{Var}(\hat{K}_i)$
0	547.849	72.907
1	557.207	95.282
2	556.211	181.048
3	530.379	119.401
4	459.725	136.771

FIGURE 2. Superimposed individual plots for the 15 mg/kg dose group.

Table 3. The values of chi-square statistics.

Hypothesis	X^2
1	0.5206
2	0.6040
3	3.8614
4	51.4473

will not eat food, and they lose weight. If we come across this kind of suspicious data, we can remove the observations that are verified as abnormal by laboratory workers and continue to analyze the remaining data set. For example, in our data set a male rate of the 200 mg/kg dose group had an abnormally large weight loss at day 91. If we delete this observation at this point, we can obtain the slightly smaller chi-square value of 50.769 in comparison with 51.447, which is given in the last row in Table 3. In this example we can conclude that the dose group of 85 mg/kg is the MNED with regard to the final body weights.

Conclusion and Discussion

We discussed the multiple comparison problems of parameter estimates assuming logistic growth-curve models jointly using a random-effect model and the closed procedure. It is important to build a model that includes only a few parameters to avoid the difficulty of handling multiplicity of the observation time points: It is similar to the problem for detecting a trend. We assumed normalities for using the random-effects model in this paper, but it is better to check the normalities of parameter estimates. This may result in devising a new technique for data analysis using random-effects models.

Some researchers might oppose using this approach to finding MNED or MED by multiple testing procedures. Suppose that no effects are observed within some dose level in mechanism A but that a small effect is observed within this dose level in mechanism B. As for mechanism B, we sometimes consider the effect not important compared to curing a disease with that drug. In addition, if we have a large enough data set, we can detect the very small effects. Statistically speaking, we do not distinguish between mechanism A and B if we do not find any significance. Therefore, we must note that the derived MNED is a statistical result and that it strongly depends on the sample size. Further research would be needed to find the MNED.

REFERENCES

1. Ruberg, S. J. Contrasts for inentifying the minimum effective dose. J. Am. Stat. Assoc. 84: 816–822 (1989).
2. Dunnett, C. W. A multiple comparison procedure fo. comparing several treatments with a control. J. Am. Stat. Assoc. 50: 1096–1121 (1955).
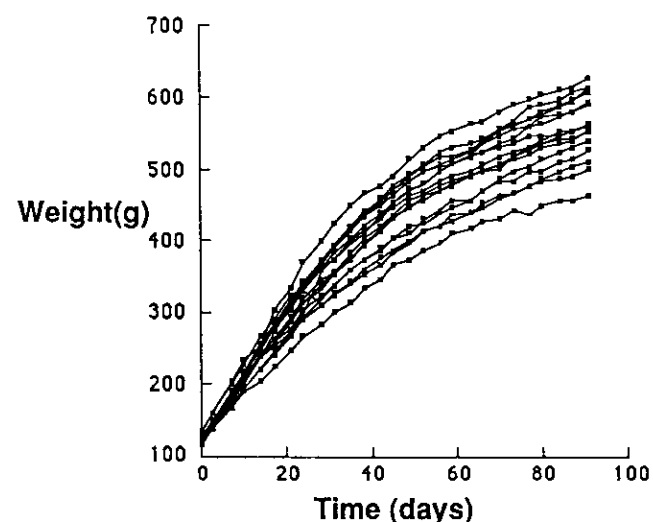
3. Williams, D. A. A test for differences between treatment means when several dose levels are compared with a zero dose control. Biometrics 28: 519–531 (1971).
4. Yoshimura, I. A study on the multiple comparisons of dose groups. In: Report of the Annual Meeting of the Japan Applied Statistical Society, 1987, pp. 47–48.
5. Korn, E., and Whittemore, A. S. Method for analyzing panel studies of acute health effects of air pollution. Biometrics 35: 795–802 (1979).
6. Marcus, R., Peritz, E., and Gabriel, K. R. On closed testing procedures with special reference to ordered analysis of variance. Biometrika 63: 655–660 (1976).